

# Somatic variant calling and interpretation in the context of cancer

**Sigve Nakken, PhD**

*Dept. of Tumor Biology*

*Institute for Cancer Research, Oslo University Hospital, Norway*

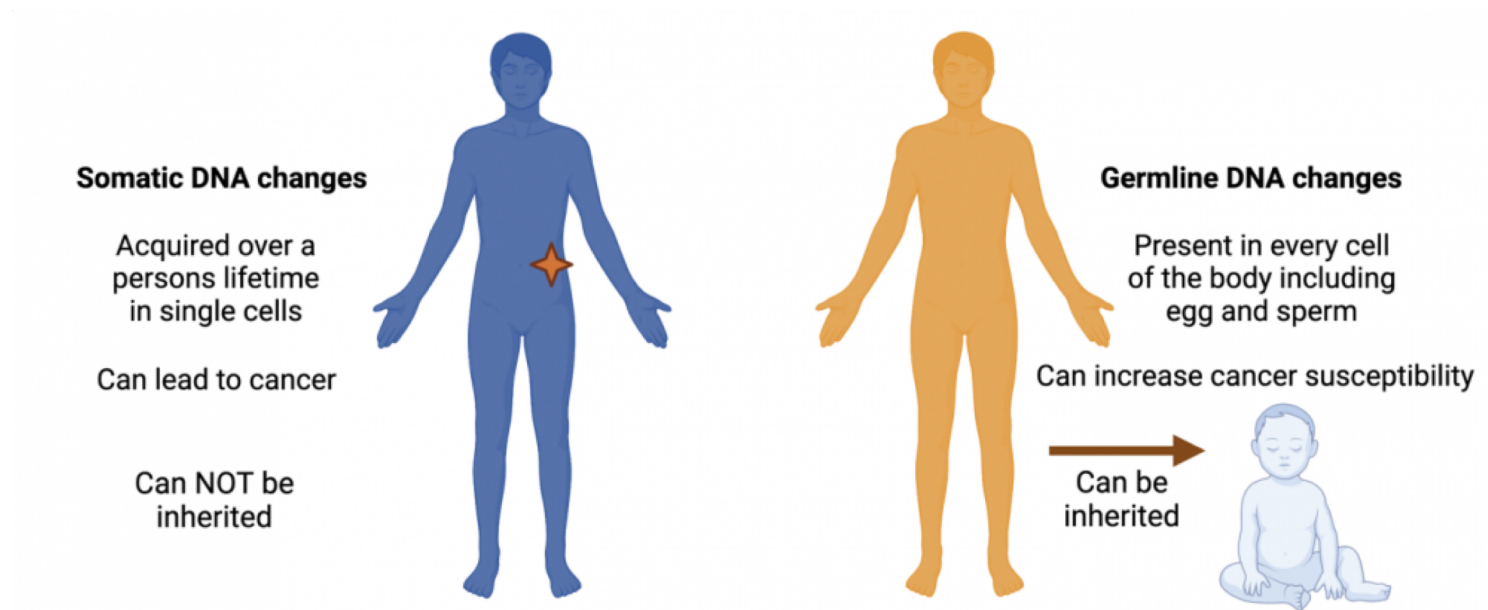
*Centre for Cancer Cell Reprogramming (CanCell), Institute of Clinical Medicine  
University of Oslo, Norway*

*Centre for Bioinformatics  
Dept. of Informatics  
University of Oslo*

# Learning outcomes – part I

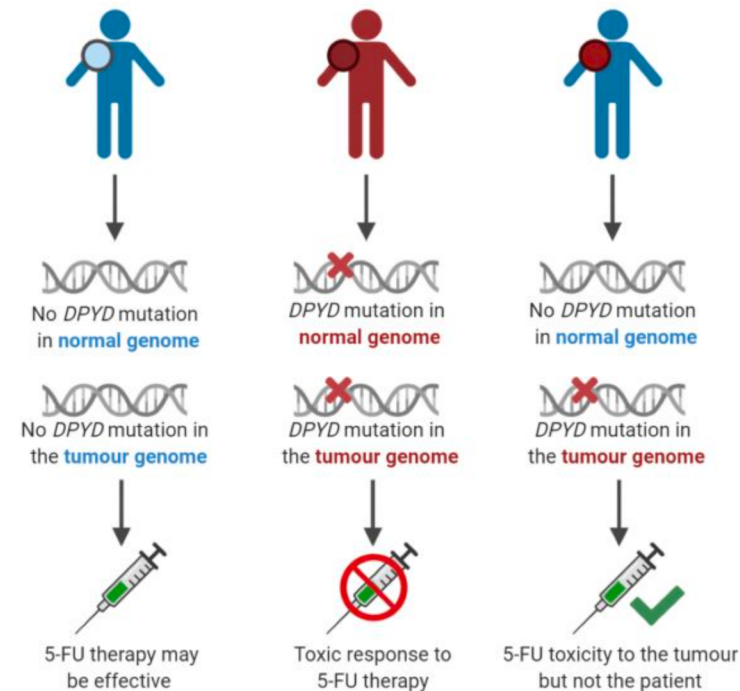
- Somatic mutations - why do we care about them?
- Which properties of tumor samples make somatic mutation calling challenging?
- Somatic variant calling principles – tumor/control vs. tumor-only
- Perspective: short-read sequencing – SNVs/InDels

# Germline versus somatic mutations



<https://www.bcgsc.ca/news/tumor-sequencing-panel-screens-both-somatic-and-germline-changes-clinical-significance>

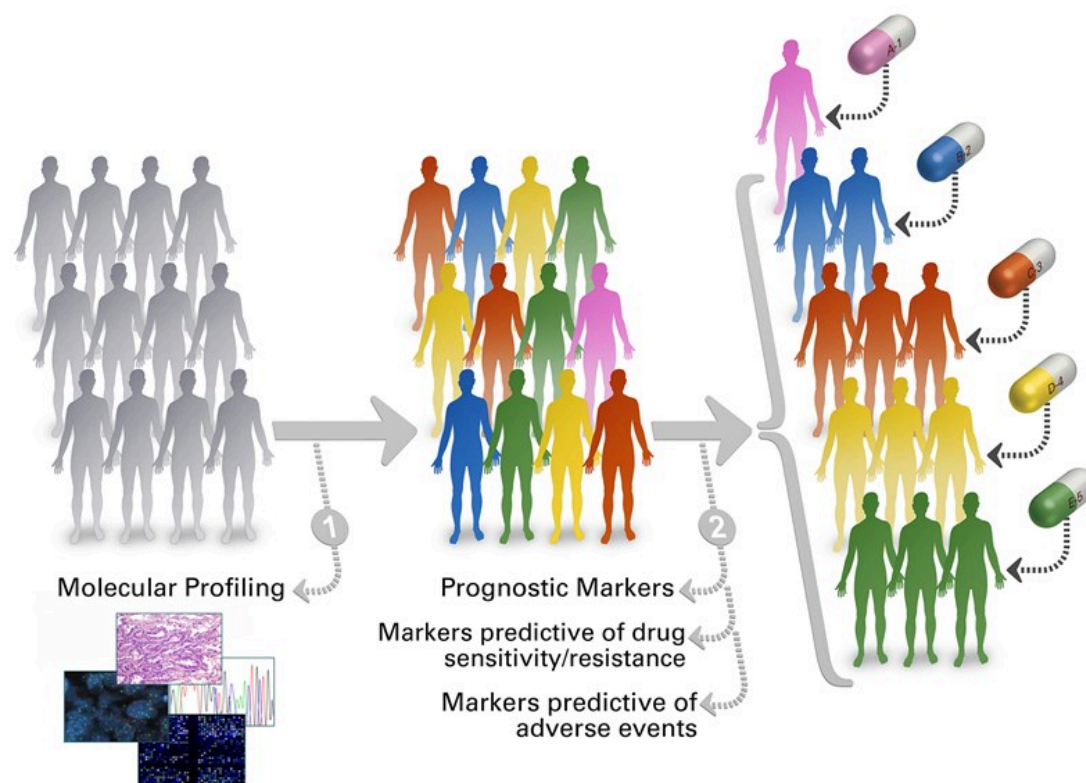
# Mutations inform upon treatment options



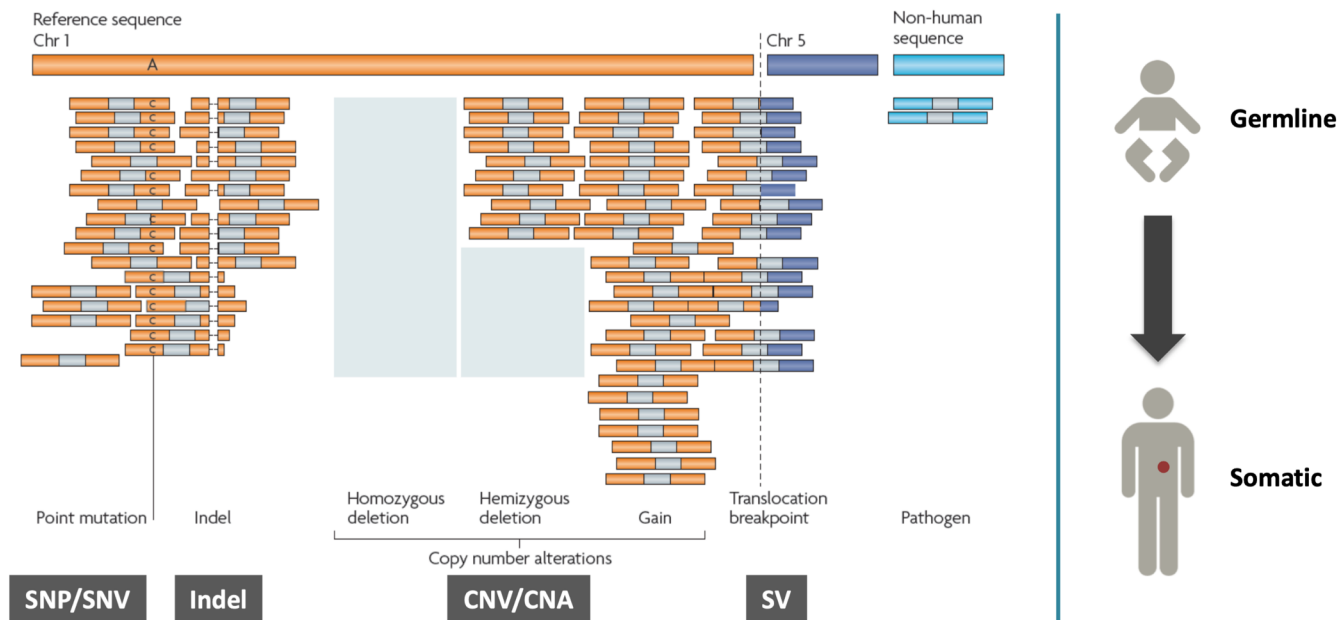
<https://www.bcgsc.ca/news/genome-sequencing-helps-prioritize-cancer-treatment-options>



# Precision cancer medicine



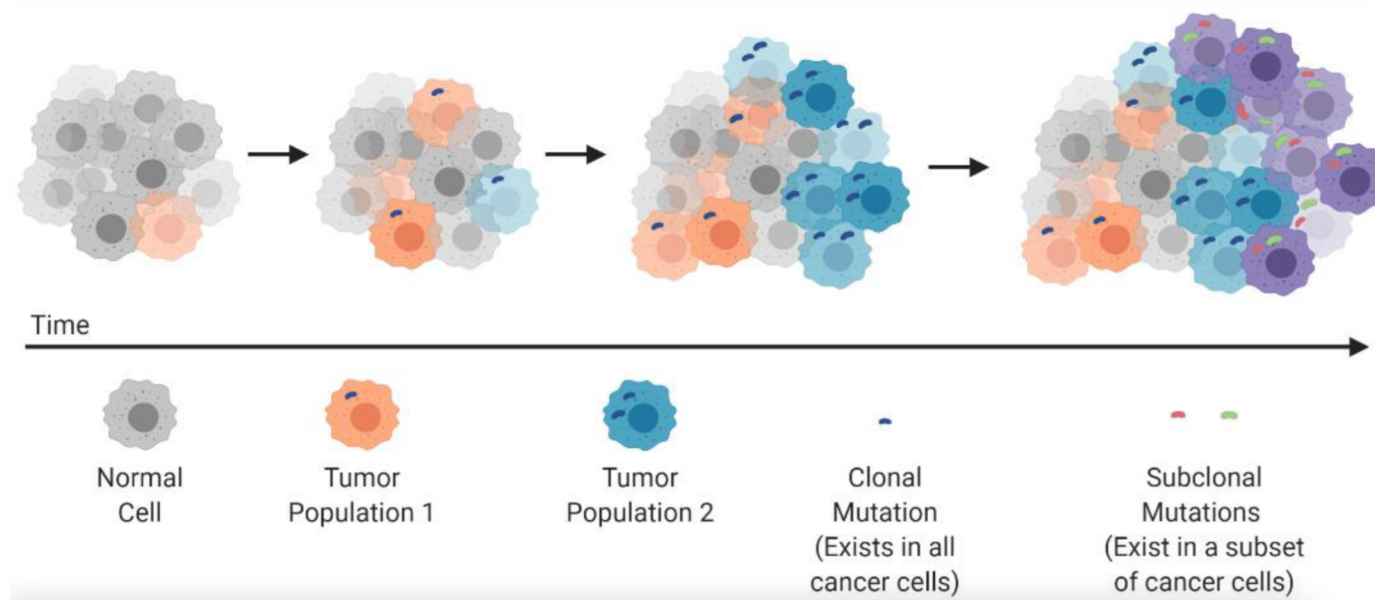
# Cancer: multiple types of DNA aberrations



GATK: Introduction to Somatic Variant Discovery

# Why is somatic variant calling so challenging?

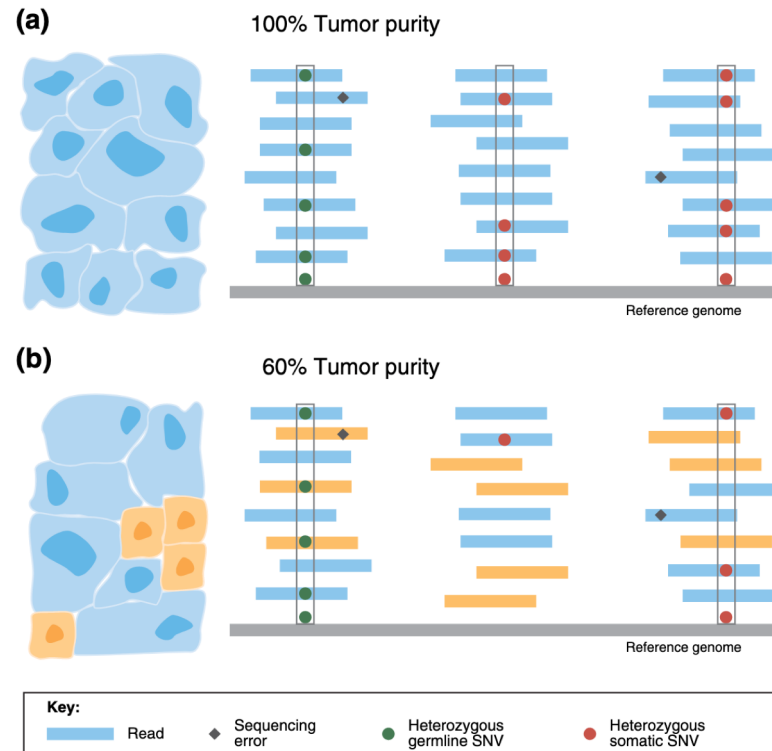
# Cancer complexity - tumor purity and heterogeneity (I)



*El-Sayes et al., Cancers, 2021*

- *Cancer sequencing provides a snapshot in time and space*

## Cancer complexity - tumor purity and ploidy (II)

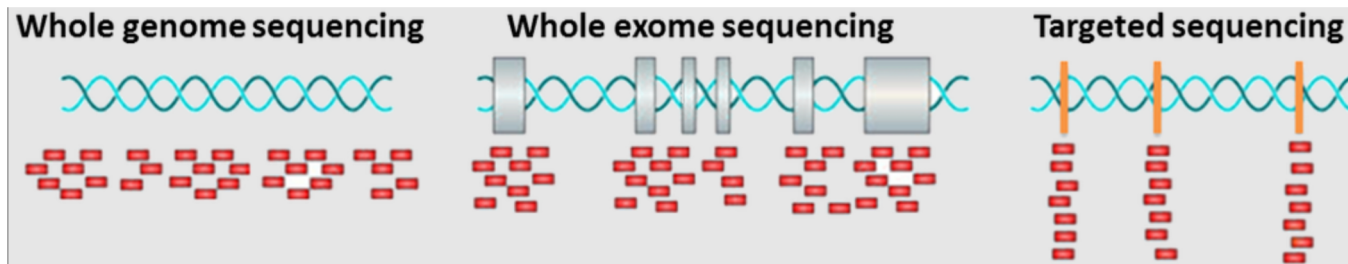


Raphael et al., Genome Med, 2014

$$\text{Tumor purity} = \frac{\text{tumor cells}}{(\text{normal} + \text{tumor cells})}$$

- Detection capability of low-frequency alleles
- Implications for targeted sequencing design
- Purity is traditionally assessed manually by pathologists, but can also be inferred computationally
- **Aneuploidy**, an imbalanced complement of chromosomes – hallmark of cancer cells

# Cancer sequencing: assay design



- Typically 30-40x coverage
- More even coverage than WES
- Covers coding and non-coding/regulatory variation
- All types of variants (reliable detection of SVs)

- Typically 80-100x coverage
- Coding regions only
- Cost-effective

- Typically > 300X coverage – captures subclonal variants at low allele frequencies
- Custom target regions – i.e. cancer-relevant/clinically actionable genes
- Most cost-effective

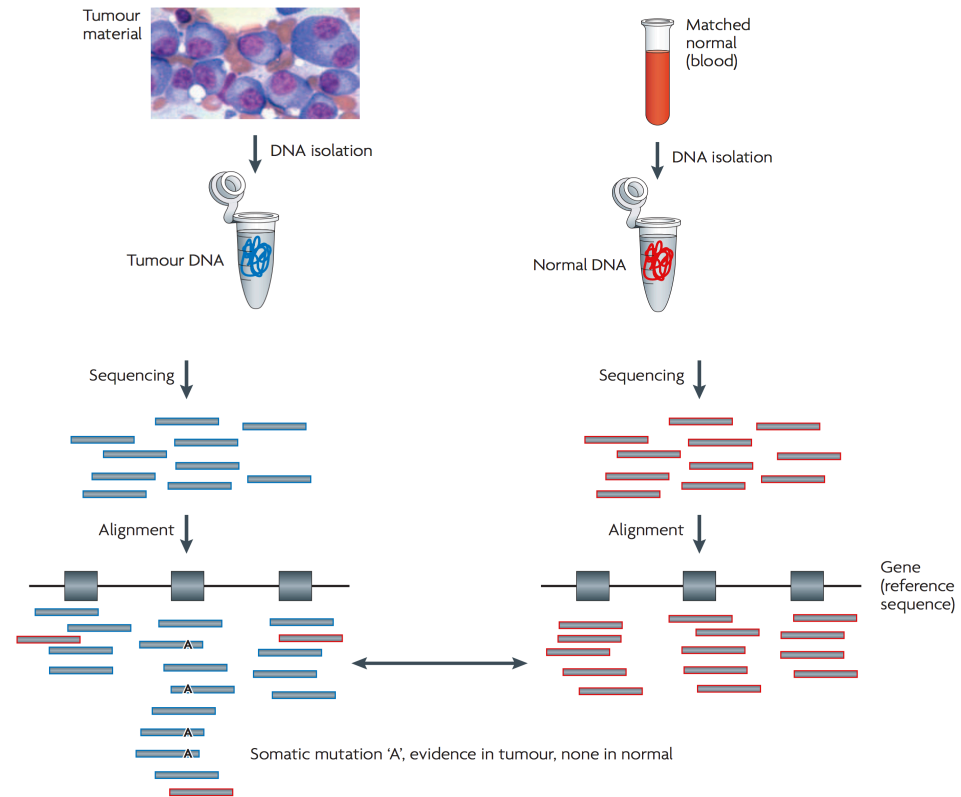
*Research*



*Clinical applications*

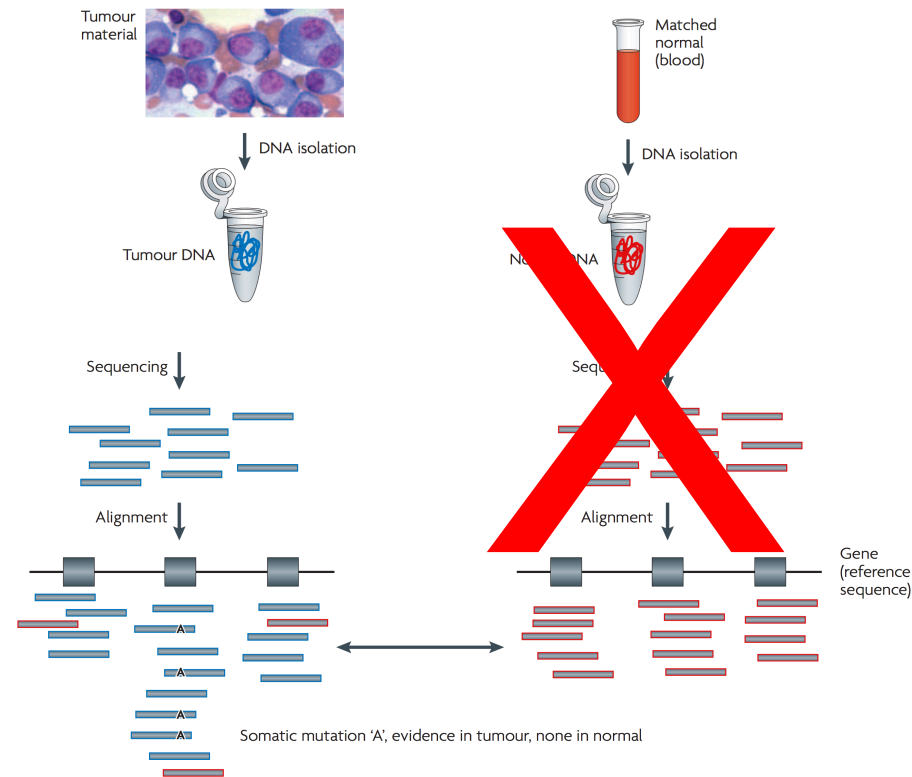
# Cancer sequencing: somatic variant calling design

- Two typical sequencing designs for detection of somatic variants
  - Tumor-control (**T + N**): most accurate



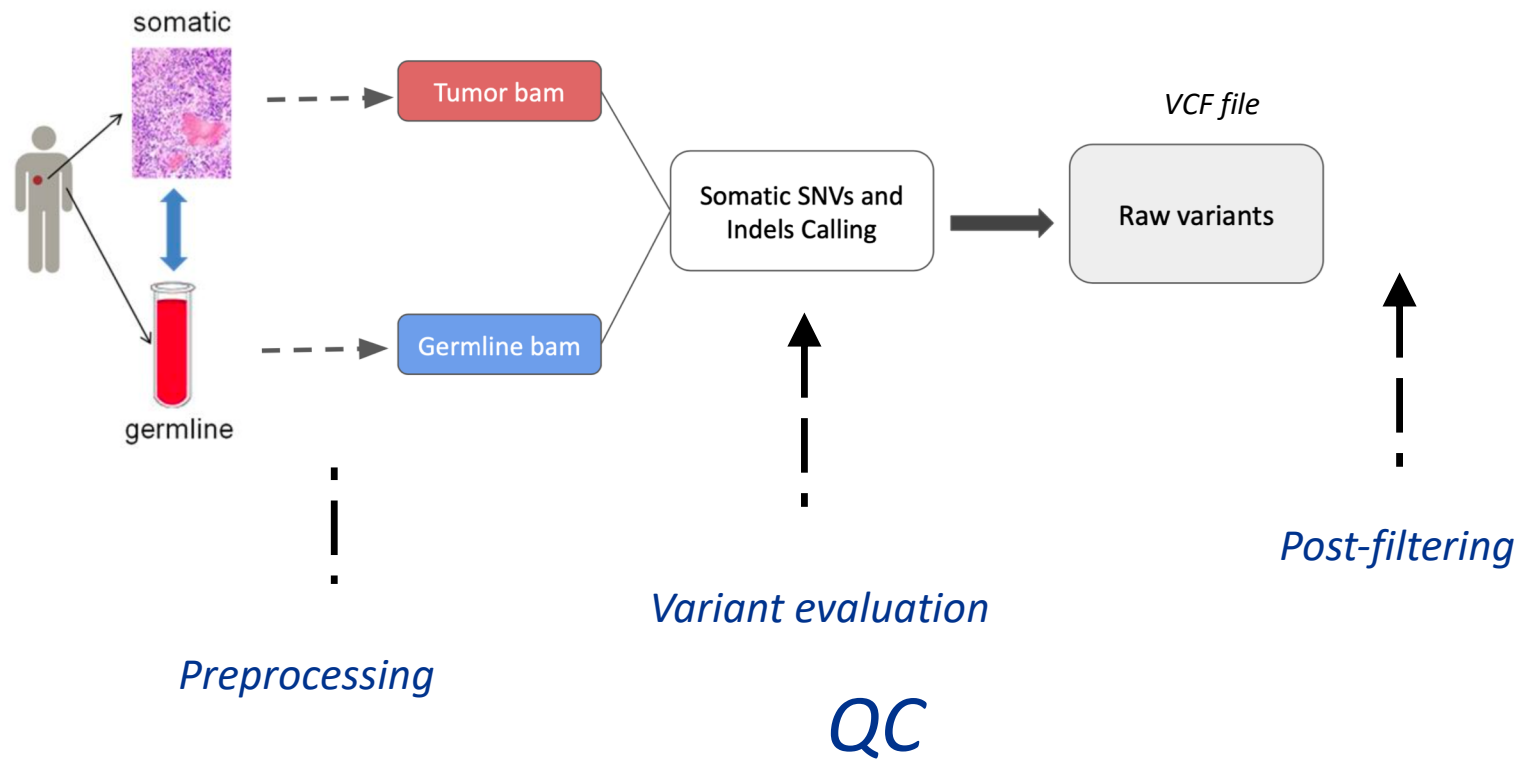
# Cancer sequencing: somatic variant calling design

- Two typical sequencing designs for detection of somatic variants
  - Tumor-control (T + N): most accurate
  - Tumor-only: most cost-effective





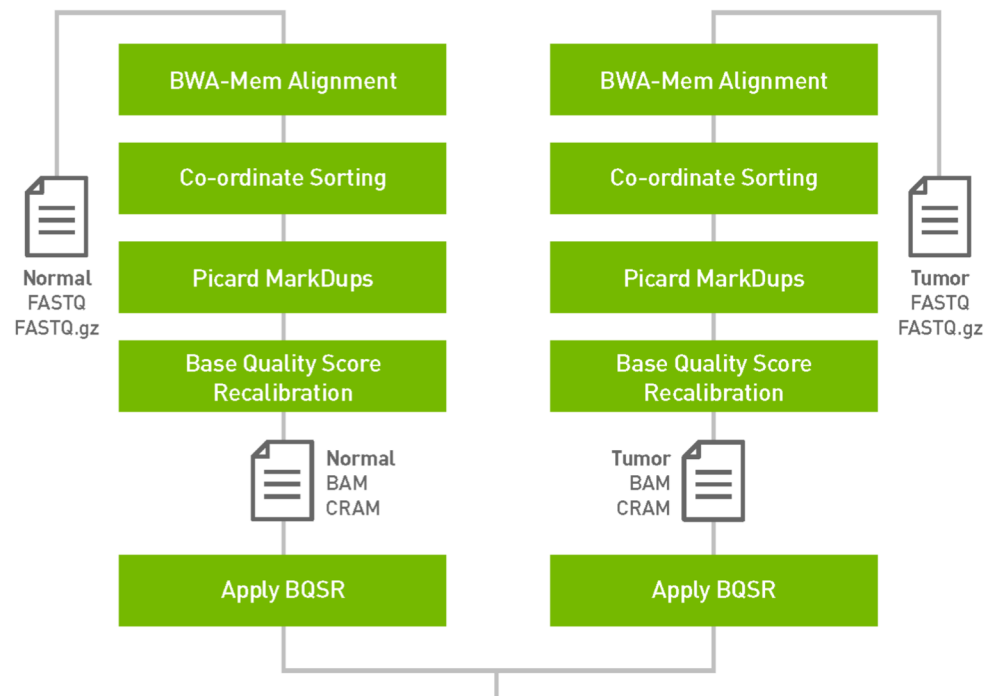
# Somatic variant calling



GATK: Introduction to Somatic Variant Discovery

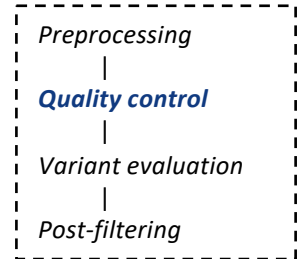
# Somatic variant calling: pre-processing

*Preprocessing*  
|  
*Quality control*  
|  
*Variant evaluation*  
|  
*Post-filtering*



<https://docs.nvidia.com/clara/parabricks/v3.0/>

# Quality control (I)

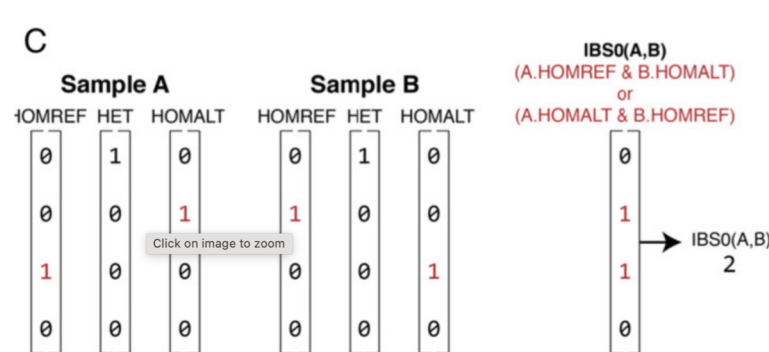


- Tumor samples subject to chemical DNA damage during sample preparation could confound variant identification
  - Oxidation-induced C>A:G>T variants
- **Detection?**
  - Imbalance between complementary nucleotide substitutions
  - Tools: **GATK**

## Quality control (II)

Preprocessing  
|  
**Quality control**  
|  
Variant evaluation  
|  
Post-filtering

- Cross-sample contamination and sample relatedness
  - Different samples are frequently handled/sequenced together
    - Cross-individual contamination may occur, even small levels of contamination will have an impact on somatic variant detection
  - **T + N:** Check that tumor and normal sample come from the same individual!
  - Tools: **Conpair/Somalier**



Pedersen et al., Genome Med, 2020

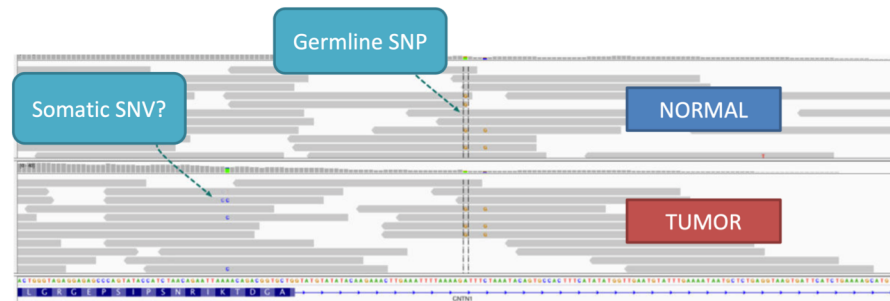
# Somatic variant calling (T + N)



Preprocessing  
|  
Quality control  
|  
**Variant evaluation**  
|  
Post-filtering

- Underlying principle for somatic variant calling algorithms using tumor-normal design: **"subtract" the germline background**
  - For a given candidate site, is the difference between tumor and normal significant?

# Somatic variant calling (T + N)



Preprocessing  
|  
Quality control  
|  
Variant evaluation  
|  
Post-filtering

- **First generation:** call somatic candidates through heuristic rules/ad-hoc filters
  - Rule out sequencing artefacts by thresholds (number of supporting reads etc)
  - Does the read count distribution support a variant (and not noise)?
  - Callers: **VarScan**, **VarDict**
- **Second generation:** probabilistic modeling of allele frequencies
  - Hypotheses/prior probabilities and likelihood estimation
  - Callers: **MuTect2**, **Strelka2**
  - **Post-filtering:** add additional quality control on call set to remove false positives (mapping quality, unbiased strand distribution etc.)

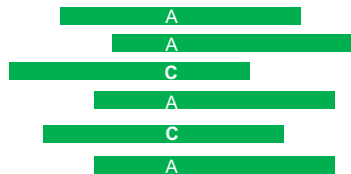
Human Reference Genome

A

Control/Normal



Tumor



# Somatic variant calling (T + N)

## MuTect2: Likelihood calculation for different scenarios

### Hypothesis 1: Somatic Variant Model

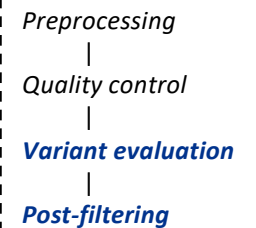
- The tumor sample has a mutation at a site that is absent in the normal sample.

### Hypothesis 2: Germline Variant Model

- A variant is present in both the tumor and normal samples (at same frequency), indicating it is inherited rather than acquired.

### Hypothesis 3: Error Model

- The observed variant in the tumor sample is due to sequencing error or other artifacts.
- The likelihood of this model depends on sequencing quality metrics like read depth, base quality, and strand bias.



Human Reference Genome

A

Control/Normal

A  
A  
A  
A  
A  
A  
A

Tumor

A  
A  
C  
A  
C  
A

## Somatic variant calling (T + N)

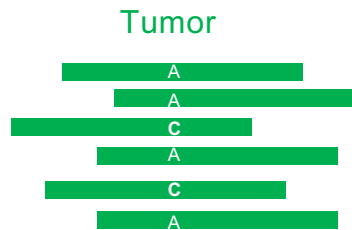
**MuTect2: choose best hypothesis through Bayesian inference**

- Prior probabilities (somatic mutations, germline variant, sequencing errors)

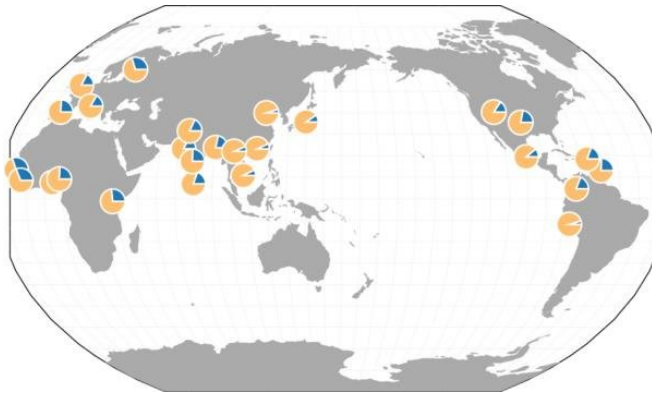
Preprocessing  
|  
Quality control  
|  
Variant evaluation  
|  
Post-filtering



# Allelic fraction versus allele frequency



***Variant allelic fraction = 2 / 6***



## ***Population allele frequency***

- *rs6602666, A>G*
- *Finnish, 0% allele frequency*
- *African 26% allele frequency*

# Somatic variant calling (T + N)

- Multiple callers exist - how to choose variant calling algorithm for a particular sequencing project?
  - Check out benchmarking results
    - A few benchmarking datasets are available – providing “gold sets” of somatic mutations
    - **Precision vs. recall**
    - Benchmarking results may be misleading
      - Which calling parameter values should be used?
  - Check whether the algorithm is designed for your assay and technology
    - E.g. has it shown good performance for detection of subclonal variants at low allele frequencies?

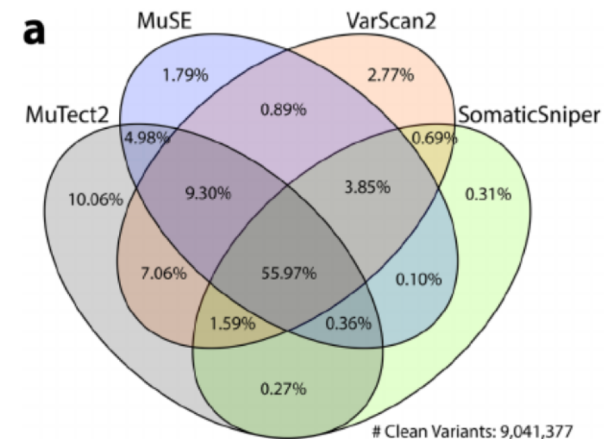


## A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing

Tyler S. Alioto, Ivo Buchhalter, Sophia Derdak, Barbara Hutter, Matthew D. Eldridge, Eivind Hovig, Lawrence E. Heisler, Timothy A. Beck, Jared T. Simpson, Laurie Tonon, Anne-Sophie Sertier, Ann-Marie Patch, Natalie Jäger, Philip Ginsbach, Ruben Drews, Nagarajan Paramasivam, Rolf Kabbe, Sasithorn Chotewutmontri, Nicole Diessi, Christopher Previti, Sabine Schmidt, Benedikt Brors, Lars Feuerbach, Michael Heinold, Susanne Gröbner, Andrey Korshunov, Patrick S. Tarpey, Adam P. Butler, Jonathan Hinton, David Jones, Andrew Menzies, Keiran Raine, Rebecca Shepherd, Lucy Stebbings, Jon W. Teague, Paolo Ribeca, Francesc Castro Giner, Sergi Beltran, Emanuele Raineri, Marc Dabad, Simon C. Heath, Marta Gut, Robert E. Denroche, Nicholas J. Harding, Takafumi N. Yamaguchi, Akihiro Fujimoto, Hideaki Nakagawa, Victor Quesada, Rafael Valdés-Mas, Sigve Nakken, Daniel Vodák, Lawrence Bower, Andrew G. Lynch, Charlotte L. Anderson, Nicola Waddell, John V. Pearson, Sean M. Grimmond, Myron Peto, Paul Spellman, Minghui He, Cyriac Kandoth, Semin Lee, John Zhang, Louis Létourneau, Singer Ma, Sahil Seth, David Torrents, Liu Xi, David A. Wheeler, Carlos López-Otin, Elias Campo, Peter J. Campbell, Paul C. Boutros, Xose S. Puente, Daniela S. Gerhard, Stefan M. Pfister, John D. McPherson, Thomas J. Hudson, Matthias Schlesner, Peter Lichter, Roland Ellis, David T. W. Jones & Ivo G. Gut - Show fewer authors

# Somatic variant calling (T + N)

- How to choose variant calling algorithm for a particular sequencing project?
  - Each caller typically has some strengths and weaknesses
    - a common strategy is to **apply multiple callers and combine the variant sets**
    - “The wisdom of crowds”
    - Union vs. intersection
      - Balance between precision and recall
    - Combining information from VCF files/callers are frequently challenging in practice



Zhang et al., Nat Commun, 2021

# Somatic variant calling: VCF

#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT CPCT02080287R CPC  
T02080287T

1 854389 . G A 590 **PASS**

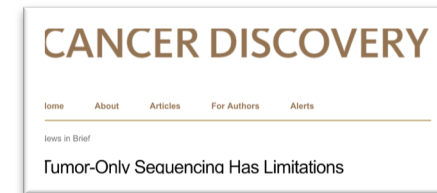
IMPACT=LINC02593,ENST00000609207,non\_coding\_transcript\_exon\_variant,  
NONE,false,n.2008C>T,,,NONE,1

GT:**AD:AF:DP**:RABQ:RAD:RC\_CNT:RC\_IPC:RC\_JIT:RC\_QUAL:RDP:SB 0/0:**42,0:0:42**:1  
528,0:42,0:0,0,0,0,42,42:0:0,0:0,0,0,0,1097,1097:42:0 0/1:**43,20:0.317:63**:1626,741:46  
,21:17,3,0,0,0,43,63:0:0,0,0:533,57,0,0,0,1318,1908:67:0.35

↑  
-  
*Allelic support – tumor  
sample*

↑  
|  
*Allelic support – normal  
sample*

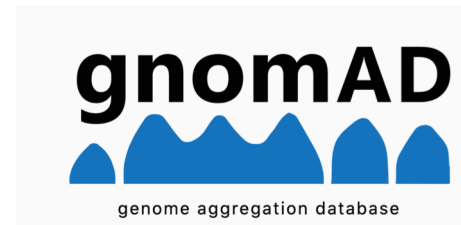
# Somatic variant calling: Tumor-only



- Cost-effective strategy for identification of somatic variants – much used in the clinic
- Main challenge: robust subtraction/filtering of the germline background
  - Approach: use other sources of germline variation (databases)
  - Each individual carries also private variants (i.e. *singletons*)
  - Ethnic subpopulations are under-represented in germline variant databases
- Targeted sequencing assays/known cancer hotspots – *a priori* assumption that findings are somatic/cancer-relevant

# Tumor-only variant filtering: gnomAD

- genome **A**ggregation **D**atabase
- **Harmonizes** germline variant data, both exome and genome sequencing, from a wide variety of large-scale sequencing projects
- **Freely available** to the scientific community
- ~735,000 WES samples
- ~75,000 WGS samples



|                        | ExAC   | gnomAD v2 | gnomAD v3 |         | gnomAD v4* |                       |
|------------------------|--------|-----------|-----------|---------|------------|-----------------------|
|                        | #      | #         | #         | #       | %          | Fold increase from v2 |
| Admixed American       | 5,789  | 17,720    | 7,647     | 30,019  | 3.72%      | 1.7x                  |
| African                | 5,203  | 12,487    | 20,744    | 37,545  | 4.65%      | 3x                    |
| Ashkenazi Jewish       | -      | 5,185     | 1,736     | 14,804  | 1.83%      | 2.9x                  |
| East Asian             | 4,327  | 9,977     | 2,604     | 22,448  | 2.78%      | 2.3x                  |
| European^              | 36,667 | 77,165    | 39,345    | 622,057 | 77.07%     | 8.1x                  |
| Middle Eastern         | -      | -         | 158       | 3,031   | 0.38%      | New                   |
| Remaining Individuals^ | 454    | 3,614     | 1,503     | 31,172  | 3.93%      | 8.8x                  |
| South Asian            | 8,256  | 15,308    | 2,419     | 45,546  | 5.64%      | 3x                    |
| Total                  | 60,706 | 141,456   | 76,156    | -       | 807,162    | -                     |

\*v4 includes all v3 samples

^ Due to small sample sizes Finnish was included in European and Amish was included in Remaining Individuals

# Tumor-only variant filtering: **norgene**

## Norwegian Germline variants browser



**Explore the Norwegian Germline variations database**

Norwegian Cancer Genomics Consortiums database of normal variation in the Norwegian population. This database currently contains 1 547 121 individual variants coming from 1590 normal chromosomes of cancer patients. Genome build hg19/GRCh37.

[Enter](#)

Based on vcf-miner from Mayo Clinic.  
The funding was provided by the [Center for Individualized Medicine](#) at Mayo Clinic.  
[Terms and Conditions of Use](#)

***norgene.no***

## Tumor-only variant filtering: **panel-of-normals**

- What is a «panel-of-normals (PON)»?
  - Variant calls made from a set of unrelated “normal” samples
- Purpose of PON?
  - Eliminate common/recurring technical artifacts
    - should use normals made using the same data generation techniques (e.g. same capture kit for exomes, same sequencing platform etc.)
  - Secondary purpose: also eliminates germline variants not called in the matched normal (or approximates the normal if none is available)

GATK: Introduction to Somatic Variant Discovery



## Somatic variant calling: summary

- The complexity of tumors pose challenges for variant identification – intratumor heterogeneity, tumor purity, ploidy
- WGS – WES – Targeted sequencing (research → clinic)
- Two fundamental sequencing designs: Tumor-control and tumor-only
- Multiple calling algorithms exist – each with strengths and weaknesses - a common strategy is to combine output from several callers
- Benchmarking results exist – can they be generalized?
- Understand the nature of your data/tumor and the priorities of the variant identification procedure when choosing a calling strategy

# Variants have been found – now what?

ACTG**C**CTACGTCTCACCGTCGACTTCAAATCG**C**TTAACCCGTACTCCCATGCTACTGC  
ATCTCGGGTTAACTCGACGTTTT**T**CATGCATGTGTGCACCCCAATATATATGCA**A**CTT  
TTGTGCACCTCTGTACGCGCGAGTTGGCACTGTCGCCCCTGTGTGCATGTGCACTGT  
CTC**T**CGCTGCACTGCCTACGTCTCACCGTCGACTTCAAATCG**C**TTAACCCGTACTCCC  
ATGCTACTGCATCTCGGGTTAACTCGACGTTTT**G**CATGCATGTGTGCACCCCAATATA  
TATGCA**A**CTTTTGTGCACCTCTGTACGCGCGAGTTGGCACTGTCGCCCCTGTGTGCA  
TGTGCACTGTCTC**T**CGAGTTTT**G**CATGCATGTGTGCACTGTGCACCTCTGTTACGTCT

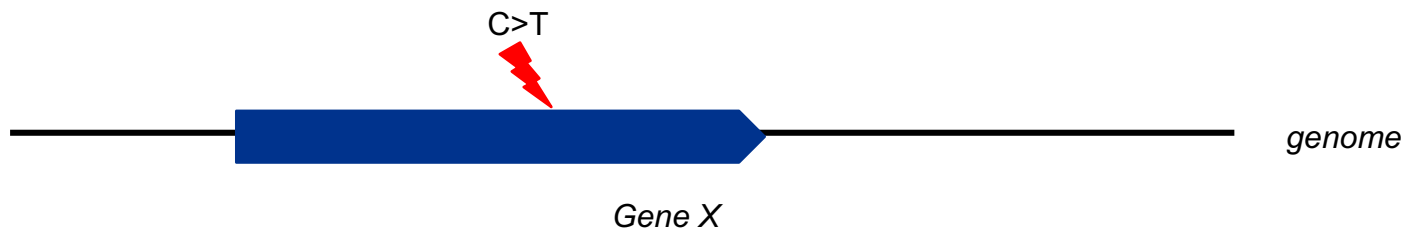


# QUESTIONS/BREAK

# Learning outcomes – part II

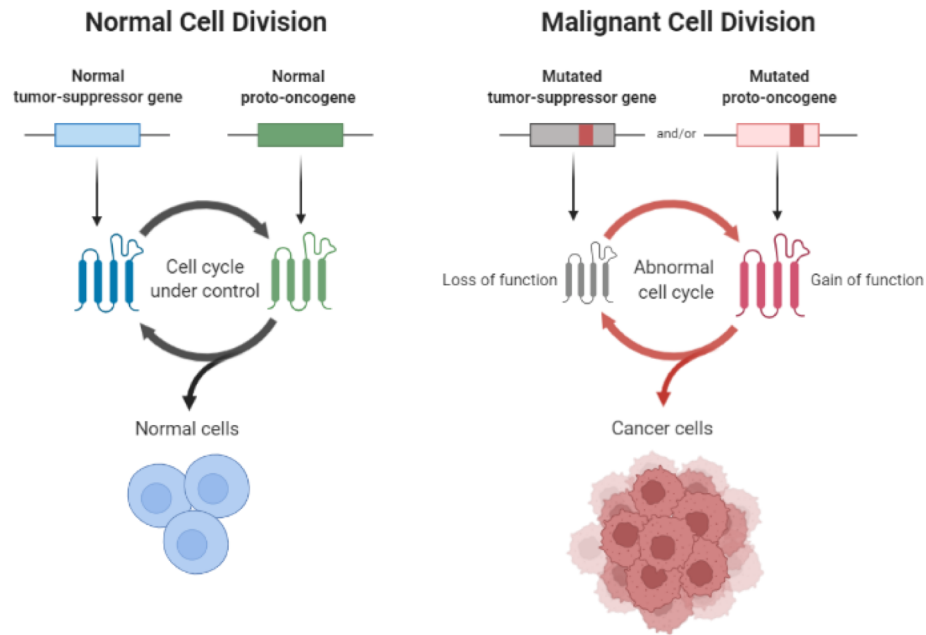
- Important cancer variant annotation concepts
  - Tumor suppressor genes versus proto-oncogenes
  - Gain-of-function versus loss-of-function variation
- Guidelines/operating procedures for variant classification and prioritization in cancer
  - Oncogenicity
  - Clinical actionability
- Mutation landscapes - «complex» biomarkers
- *Perspective – precision medicine/single sample analysis*

# Individual variant interpretation – key questions



1. Which gene is mutated?
  - Is the gene cancer-relevant?
2. What is the predicted consequence for the encoded protein?
  - Loss-of-function?
  - Gain-of-function?
3. Is it actionable?

# Which genes are mutated (II)?



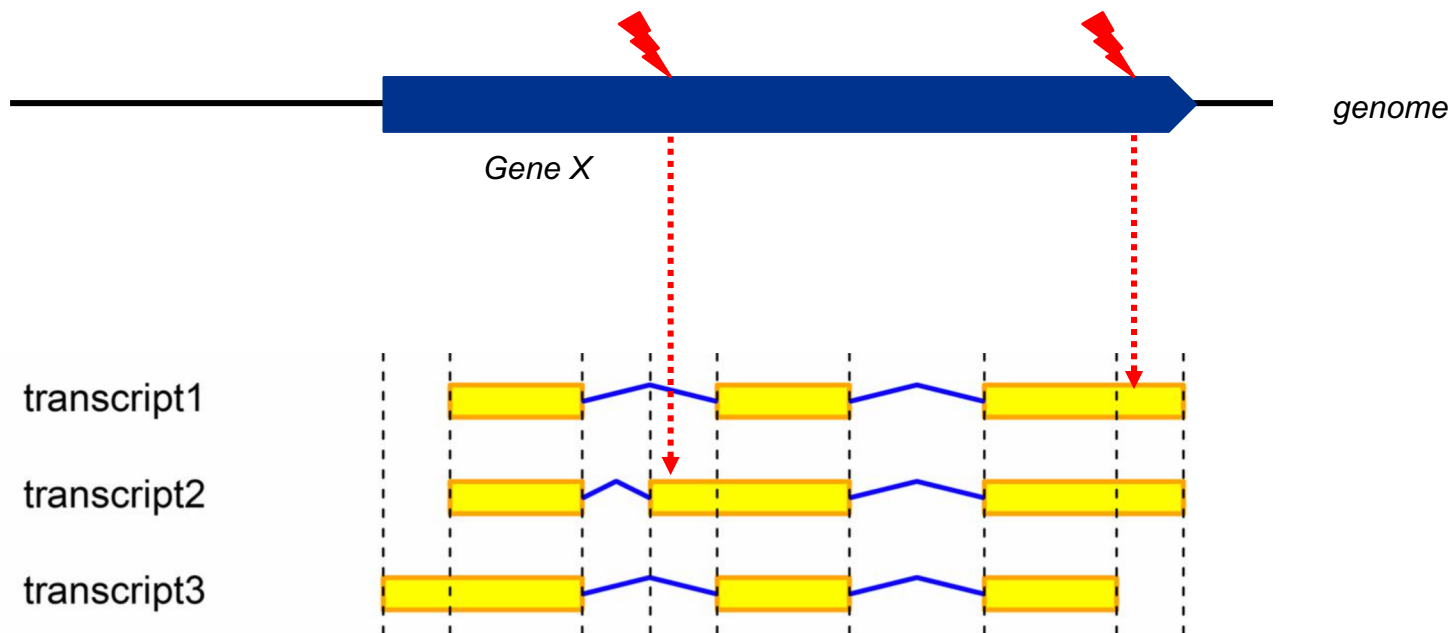
- N = 360 proto-oncogenes
- N = 372 tumor suppressor genes

## Cancer Gene Census

<https://cancer.sanger.ac.uk/census>

# Variant consequence interpretation - general

- Important: a gene typically consists of multiple transcript isoforms

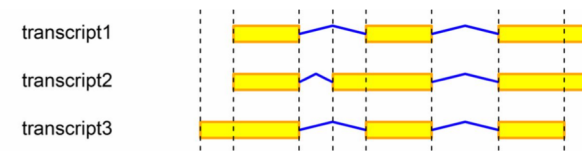


## Variant consequence interpretation – general (II)

- Several transcript/gene model databases

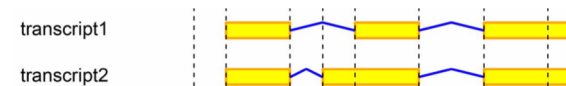
- RefSeq
- Ensembl
- GENCODE
  - MANE

*RefSeq*



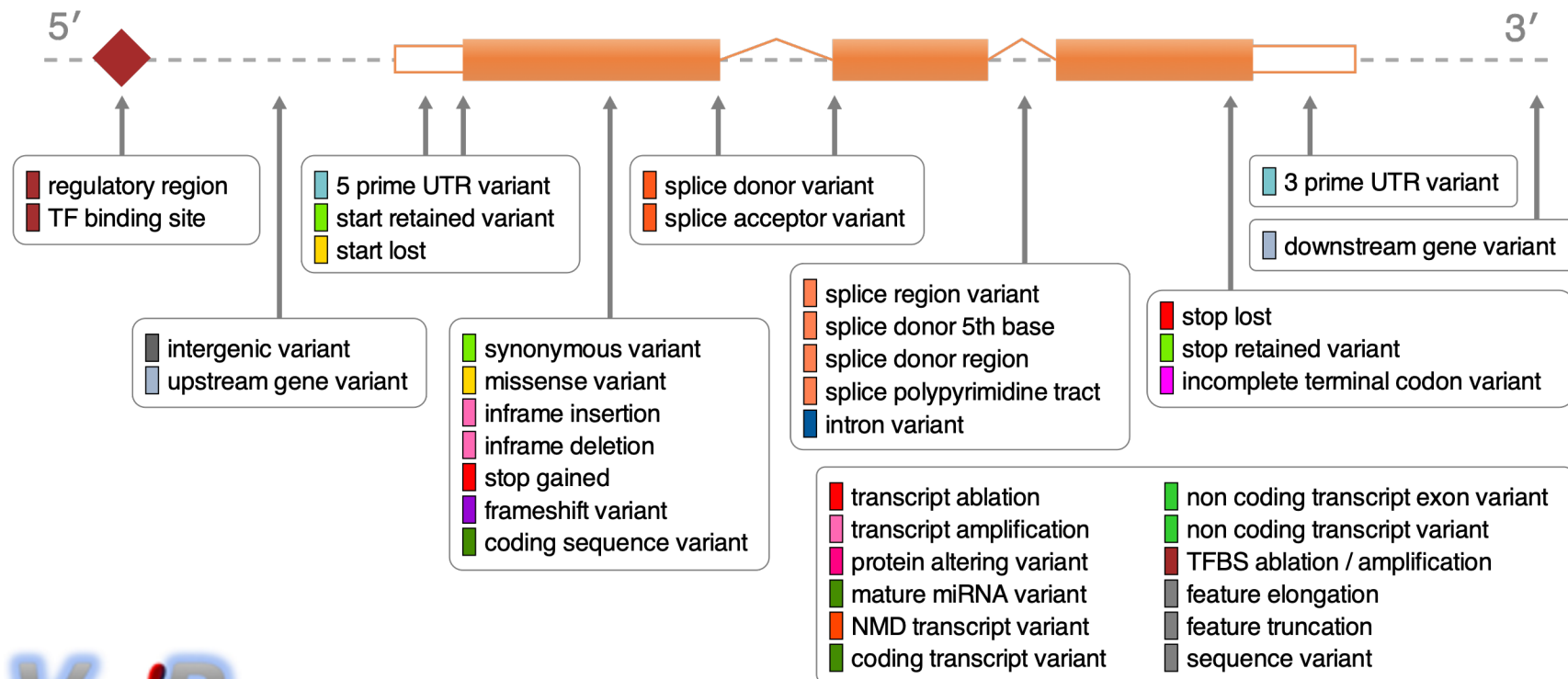
- Choice of transcript database impacts variant consequence/ annotation
- Frequent strategy: Report variant consequence in most commonly expressed isoform (i.e. ***principal*** isoform)

*GENCODE*



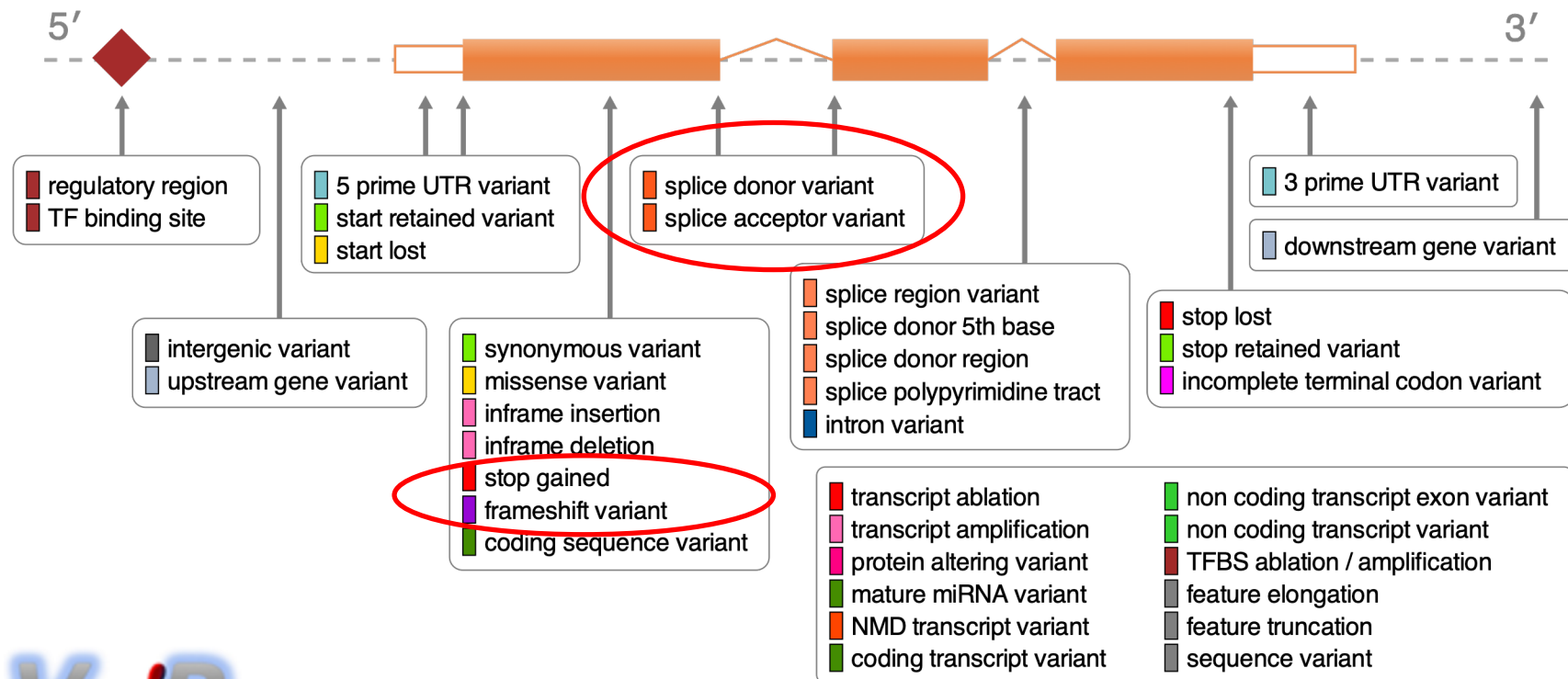


# Variant consequence interpretation – general (III)



Variant Effect Predictor

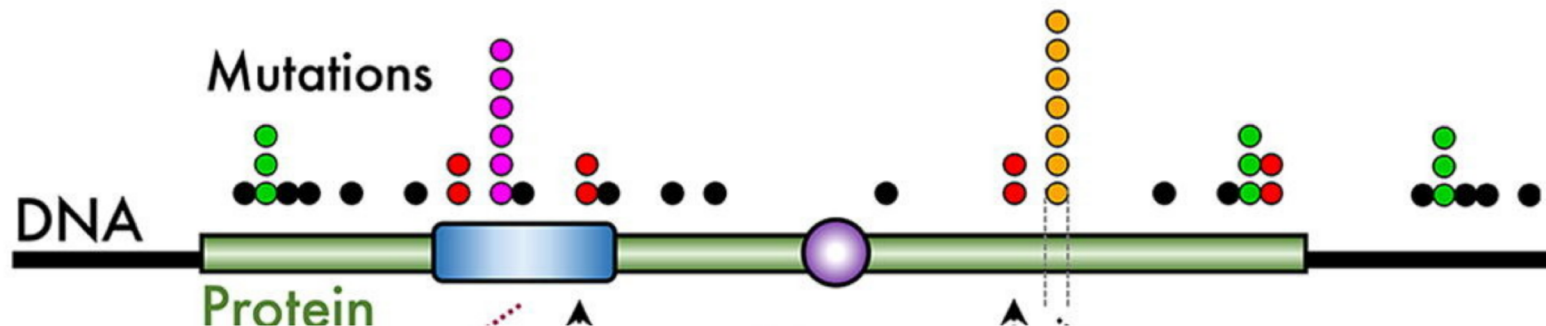
# Loss-of-function variants



Variant Effect Predictor

# Gain-of-function variants

- Much harder to predict than loss-of-function variants
- Large-scale sequencing efforts have found clusters of missense mutations in oncogenes



[cancerhotspots.org](http://cancerhotspots.org)

# Variant oncogenicity classification

- Which tumor variants are likely to be oncogenic?
  - Loss-of-function in tumor suppressor genes, gain-of-function of oncogenes
  - Guidelines/standard operating procedures for classification published

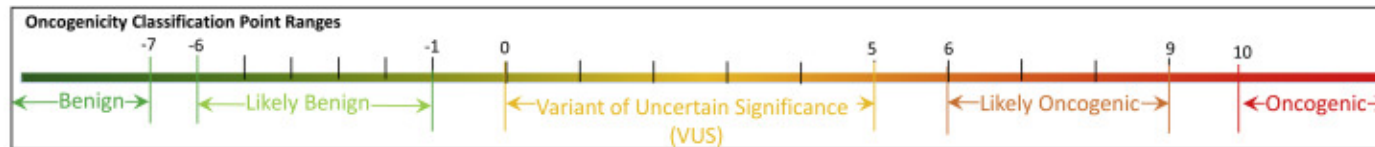
SPECIAL ARTICLE | VOLUME 24, ISSUE 5, P986-998, MAY 2022 [Download Full Issue](#)

Standards for the classification of pathogenicity of somatic variants in cancer (oncogenicity): Joint recommendations of Clinical Genome Resource (ClinGen), Cancer Genomics Consortium (CGC), and Variant Interpretation for Cancer Consortium (VICC)

Peter Horak   • Malachi Griffith • Arpad M. Danos • ... Obi L. Griffith • Debyani Chakravarty   • Dmitry Sonkin   • [Show all authors](#)

[Open Archive](#) • Published: January 28, 2022 • DOI: <https://doi.org/10.1016/j.jim.2022.01.001> •  Check for updates

# Variant oncogenicity classification

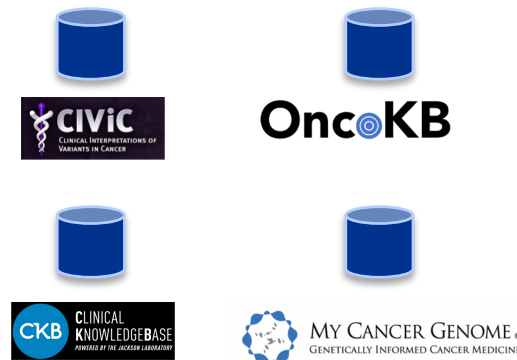


|                        | Benign      |                                                               |                                                                             | Oncogenic                                                         |                                                                                                                           |                                                                       |                                  |
|------------------------|-------------|---------------------------------------------------------------|-----------------------------------------------------------------------------|-------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------|----------------------------------|
| Evidence Strength      | Very Strong | Strong                                                        | Supporting                                                                  | Supporting                                                        | Moderate                                                                                                                  | Strong                                                                | Very Strong                      |
| POINTS                 | -8          | -4                                                            | -1                                                                          | +1                                                                | +2                                                                                                                        | +4                                                                    | +8                               |
| Population Data        | MAF is >5%  | MAF is >1%                                                    |                                                                             | Absent in population databases                                    |                                                                                                                           |                                                                       |                                  |
| Functional Data        |             | Well-established functional studies show no oncogenic effects |                                                                             |                                                                   |                                                                                                                           | Well-established functional studies supportive of an oncogenic effect |                                  |
| Predictive Data        |             |                                                               | Silent mutation (no predicted impact on splicing)                           |                                                                   | Missense change at an amino acid residue where a different missense change determined to be oncogenic has been documented | Same amino acid change as a previously established oncogenic mutation | Null variant in tumor suppressor |
| Cancer Hotspots        |             |                                                               |                                                                             | Cancer hotspots with low frequency of recurrence                  | Cancer hotspot with moderate frequency of recurrence                                                                      | Cancer hotspot with high frequency of recurrence                      |                                  |
| Computational Evidence |             |                                                               | All utilized lines of computational evidence suggest no impact of a variant | All utilized lines of computational evidence support oncogenicity |                                                                                                                           |                                                                       |                                  |

# Which variants are actionable? (I)

- Specific genetic aberrations indicate clinical actionability
  - **Drug sensitivity**
  - Prognosis / Diagnosis
  - Drug resistance
- Multiple initiatives curate clinical variant associations in cancer
  - **Variant X** in **phenotype Y** indicates sensitivity to **drug Z**
  - **Challenge:** harmonization of knowledge databases
    - VICC (Variant Interpretation for Cancer Consortium)

- *HER2 amplification – Trastuzumab – Breast Cancer*
- *BRAF V600E – Vemurafenib – Melanoma*
- *BRAF V600E – Trametinib + Dabrafenib – NSCLC*
- *IDH1/2 mutations – Ivosidenib – AML*
- ..
- ..



# Which variants are actionable? (II)

SPECIAL ARTICLE | VOLUME 29, ISSUE 9, P1895-1902, SEPTEMBER 01, 2018

A framework to rank genomic alterations as targets for cancer precision medicine: the ESMO Scale for Clinical Actionability of molecular Targets (ESCAT)

J. Mateo • D. Chakravarty • R. Dienstmann • S. Jezdic • A. Gonzalez-Perez • N. Lopez-Bigas • C.K.Y. Ng • P.L. Bedard • G. Tortora • J.-Y. Douillard • E.M. Van Allen • N. Schultz • C. Swanton • F. André • L. Pusztai • Show less

SPECIAL ARTICLE | VOLUME 19, ISSUE 1, P4-23, JANUARY 01, 2017

Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer

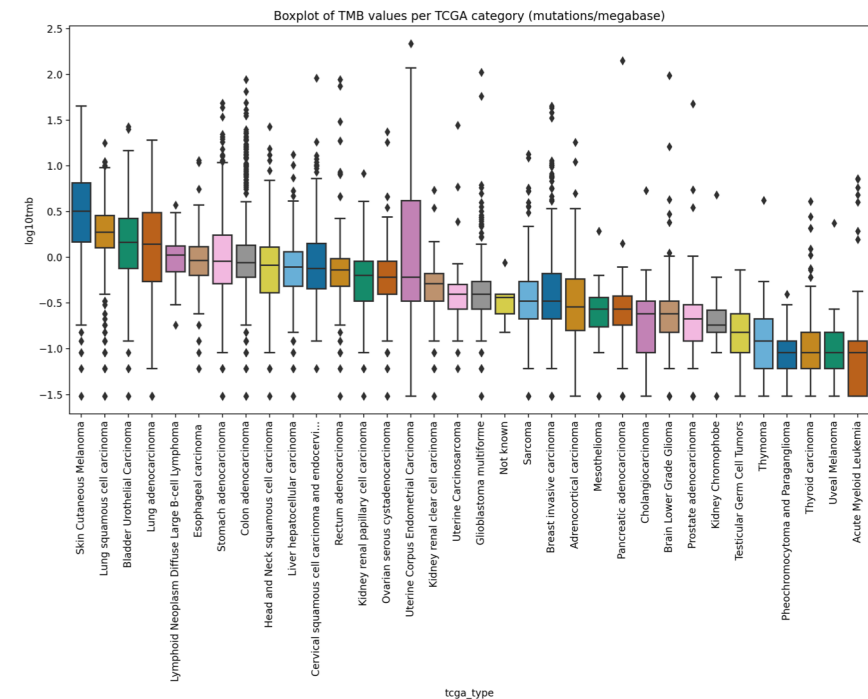
A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists

Marilyn M. Li • Michael Datto • Eric J. Duncavage • Shashikant Kulkarni • Neal I. Lindeman • Somak Roy • Apostolia M. Tsimberidou • Cindy L. Vnencak-Jones • Daynna J. Wolff • Anas Younes • Marina N. Nikiforova • Show less

- How do we prioritize variants according to actionability?
  - Classification guidelines - tiers
  - Key: **Strength of evidence**
  - Tumor type (**on-label** vs. **off-label**)
- **TIER 1** – strong evidence for clinical impact, same tumor type as query
- **TIER 2** – strong evidence for clinical impact in other tumor type or weak evidence for clinical impact in query tumor type
- **TIER 3** – uncertain clinical significance; coding variants in tumor suppressor genes/proto-oncogenes (mutation hotspots etc)
- **TIER 4** – other coding variants

# Mutation landscapes – complex biomarkers

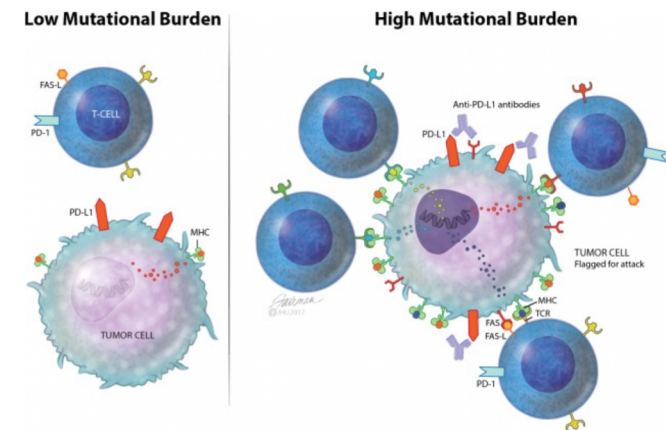
- Properties of tumor samples not linked to individual variants but rather the global pattern or landscape of mutations found
1. Tumor mutational burden
  2. Mutational signatures
  3. Microsatellite instability
    - Important in selected tumor types





# Mutational burden - how many tumor mutations are found?

- **Tumor mutational burden (TMB)** - number of somatic mutations per megabase of interrogated genomic sequence
- A key driver in the generation of immunogenic neoepitopes – influences **response to immune checkpoint inhibitors (ICIs)**
- Clinically significant TMB levels - debatable
- Very simple biomarker – still a main molecular indicator for allocation of patients to immunotherapy



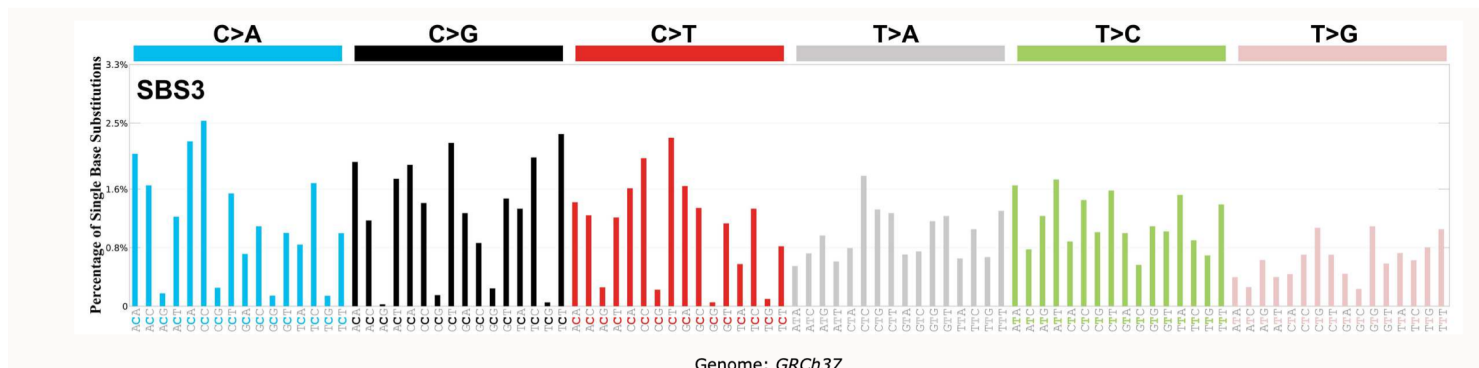
*Illustration from Sharabi et al., The Oncologist (2017)*

# What types of mutations are found?

- **Mutational signatures:** characteristic mutation patterns (types and sequence context) that arise from a specific mutational process
- **Premise:** mutational processes are context-dependent (occur non-randomly in DNA)
- **Footprint:** The global set of mutations harvested from NGS **reveals a «historical footprint»** of the mutational processes that have shaped a given tumor
  - Environmental mutagens
  - Endogenous mutation processes (e.g. DNA repair defects)
  - Treatment effects
  - **Approximately 60 established mutational signatures**

# Mutational signatures (I)

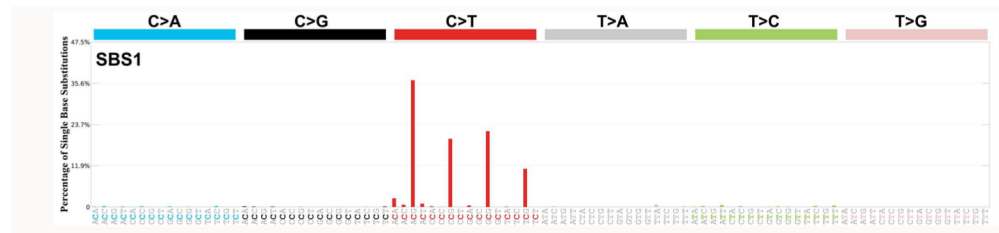
- Mutational signatures are most commonly presented through the **96-channel** approach (single base substitutions, SBS)
  - Mutation type + flanking bases



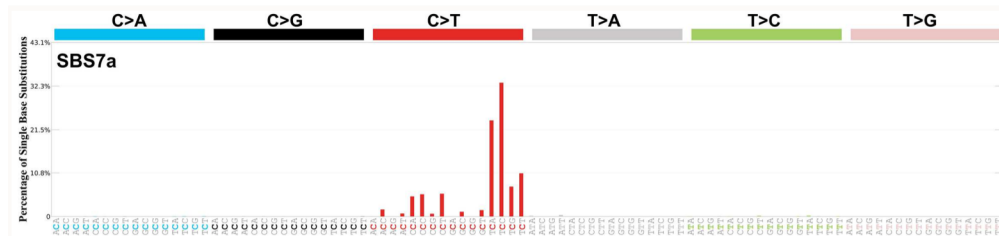
- A single signature (attributed to a given process) is thus characterized as the **relative frequency of 96 different channels**

## Mutational signatures (II)

- Aging
  - spontaneous or enzymatic deamination of 5-methylcytosine to thymine (clock-like signature)

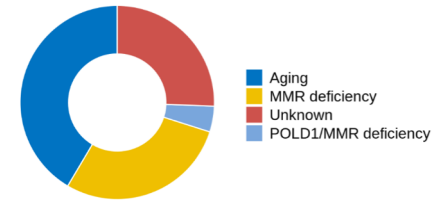


- Exposure to UV light
  - cyclobutane pyrimidine dimers or 6-4 photoproducts



# Mutational signatures (III)

- Tools can «deconstruct» the profile of somatic mutations in a tumor towards contribution of known signatures
- Signatures are emerging as an important biomarker for drug response
- Often considered in combination with other markers
- Challenge: **confidence**

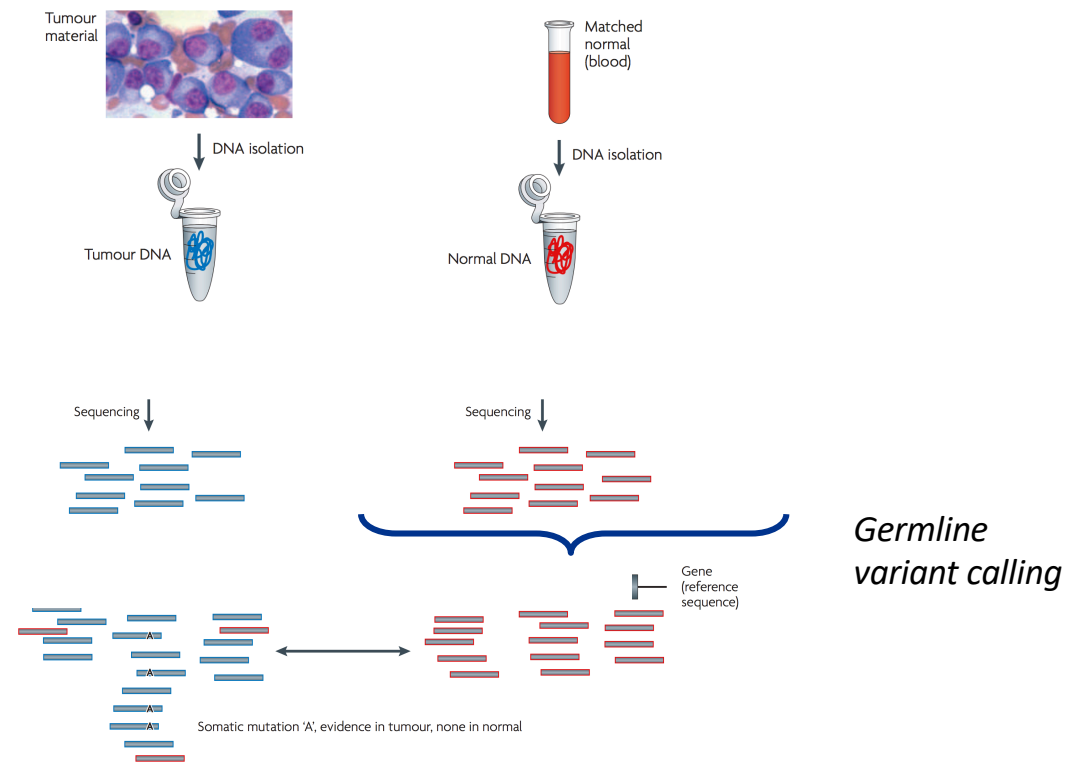


|      |       |       |       |                                                |                                                                                    |
|------|-------|-------|-------|------------------------------------------------|------------------------------------------------------------------------------------|
|      |       | CS-3  | CS-8  | Homologous Recombination Repair Deficiency     | PARP inhibition <sup>32-34</sup> ,<br>Platinum-based chemotherapy <sup>35-37</sup> |
| CS-6 | CS-15 | CS-20 | CS-26 | Mismatch Repair Deficiency                     | PD1-immunotherapy <sup>48-49,52</sup>                                              |
|      |       | CS-5  | CS-8  | Nucleotide Excision Repair Deficiency          | Cisplatin <sup>63-65</sup>                                                         |
|      |       | CS-18 | CS-30 | Base excision Repair Deficiency                |                                                                                    |
|      |       |       | CS-10 | Deficient DNA polymerase proofreading activity | PD1-immunotherapy <sup>48-49,52</sup>                                              |

Adopted and modified from Van Hoeck et al., BMC Cancer, 2019

# Cancer patients - germline background (I)

- Genetic causes of cancer include both somatic mutations and inherited germline variants.
- Approximately 10-15% of all cancers have a hereditary/germline component
- Pathogenic variants in cancer predisposition genes – conferring increased risk of tumor development



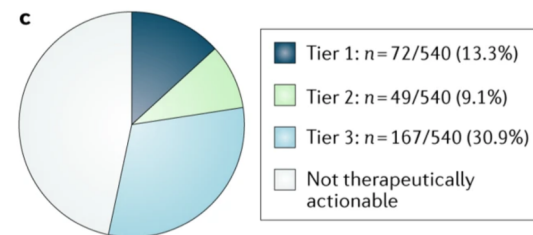
## Cancer patients - germline background (II)

- Why important?
  - Implement surveillance and risk-reducing interventions
  - May impact type of surgery (radical /conservative)
  - Targeted therapy implications
    - BRCA (PARP)



Review Article | Published: 19 February 2019

### Therapeutic implications of germline genetic findings in cancer

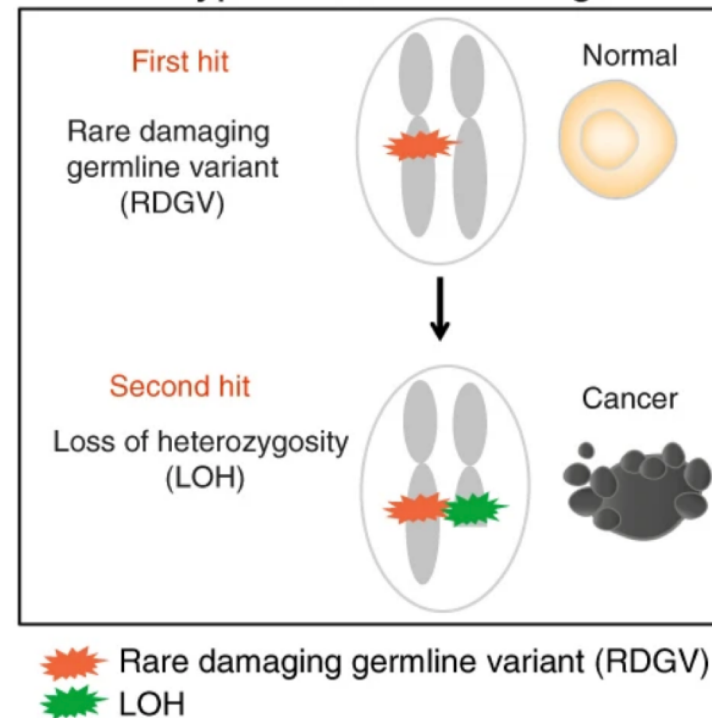


*Clinical actionability - TCGA*

## Cancer patients - germline background (III)

- Knudsen's two-hit hypothesis
  - First formulated in 1971
  - Many tumor suppressor genes require both alleles to be inactivated
  - First mutation inherited, second mutation acquired

Two-hit hypothesis for tumorigenesis





# Putting it all together

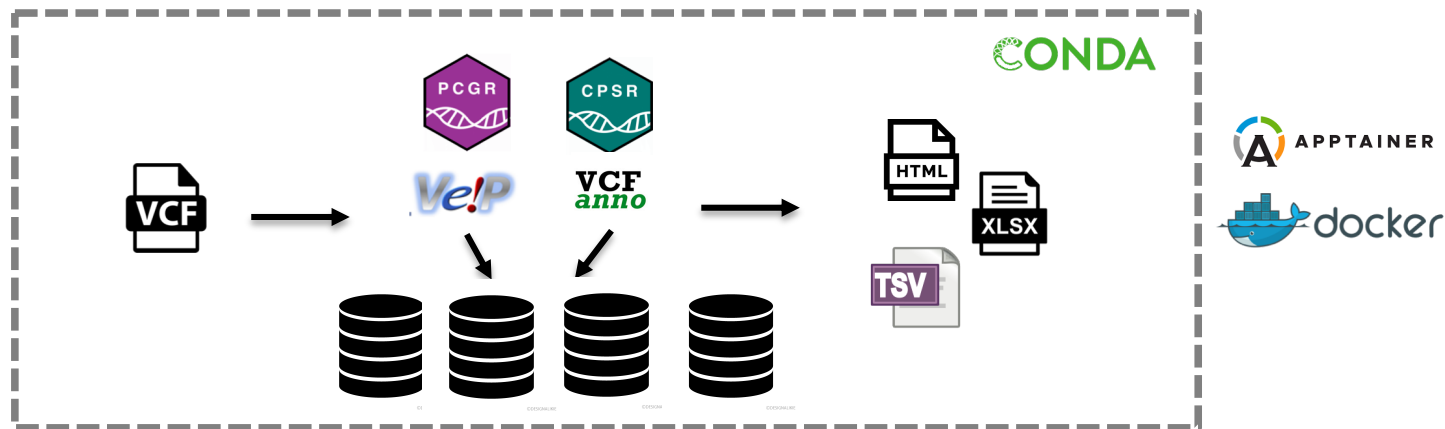
- **Variant interpretation tools for molecular-based cancer treatment**

- Personal Cancer Genome Reporter (PCGR): Translation of tumor omics data to measures for clinical decision making
  - Cancer Predisposition Sequencing Reporter (CPSR): Assessment of cancer-predisposing germline variants
- N-of-one interpretation challenge
  - What's characteristic of tumor X/patient X? Given the current knowledge/reference data
- Backbone for variant analysis in national precision medicine initiatives



## PCGR/CPSR

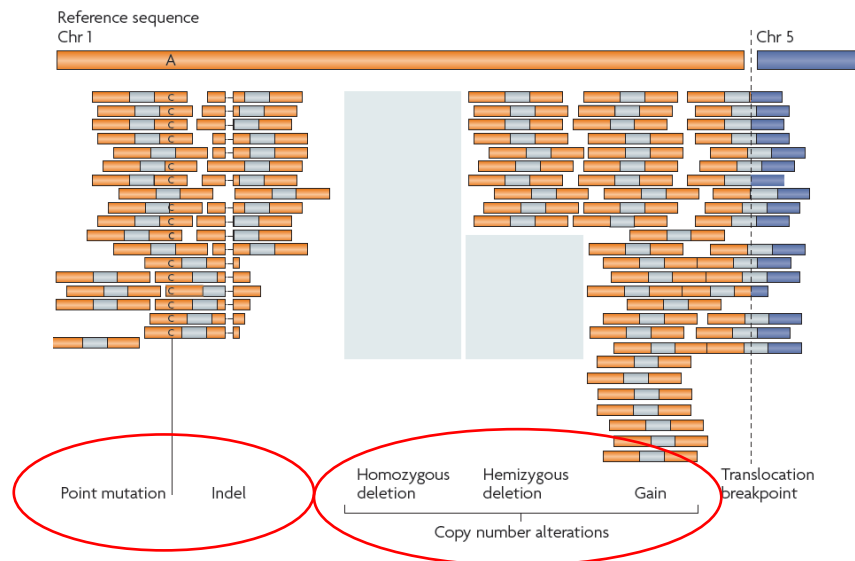
- Overall aim: **harvest precision therapy measures** from high-throughput sequencing data of cancer patients (tumor + germline)
- Focus: variant prioritization/classification
- **Stand-alone reporting engines** – **free/open-source** - publicly available reference data
- Reports on a **per-case** basis



# PCGR - input data

- Key inputs:

## 1) Somatic DNA aberrations



## 2) Tumor type

Breast  
Colon/Rectum  
Lung  
Pancreas  
Prostate  
...

## 3) Expression

## 4) Germline findings

## 5) RNA fusion events (in progress)

Meyerson et al. Nat Rev Genet 2010

# PCGR/CPSR requires prior knowledge

- What type of variant?

- Which variant consequence?
- Seen previously? Germline/somatic?  
At what frequency? In which tumor types?
- Loss-of-function?
- Gain-of-function/hotspot? In functional domain?
- Clinically actionable? Drug sensitivity/resistance marker?



- What type of gene?

- Cancer-relevant? Cancer-predisposing? Known tumor-suppressive or proto-oncogenic role? Part of oncogenic signaling pathways?
- Molecular drug target? Approved/Late phase/early phase?





# PCGR - key functionality

- **A prioritization of variants (SNVs/InDels/CNAs) with respect to functional and clinical impact**
  - Therapeutic actionability/prognosis/diagnosis
  - Oncogenic potential
- Other measures that can guide/inform upon precision treatment
  - **Mutational signatures** – what type of mutational processes have shaped the tumor
  - **MSI classification** – predicted from mutational profile
  - **Tumor mutational burden** – proxy for response to immunotherapy
  - **Germline findings**
  - **RNA expression outliers / RNA similarity analysis**
- Configurable, interactive, and transparent output
- Support for tumor-only and tumor-control input assays
- WES/WGS/targeted assays (TSO500)

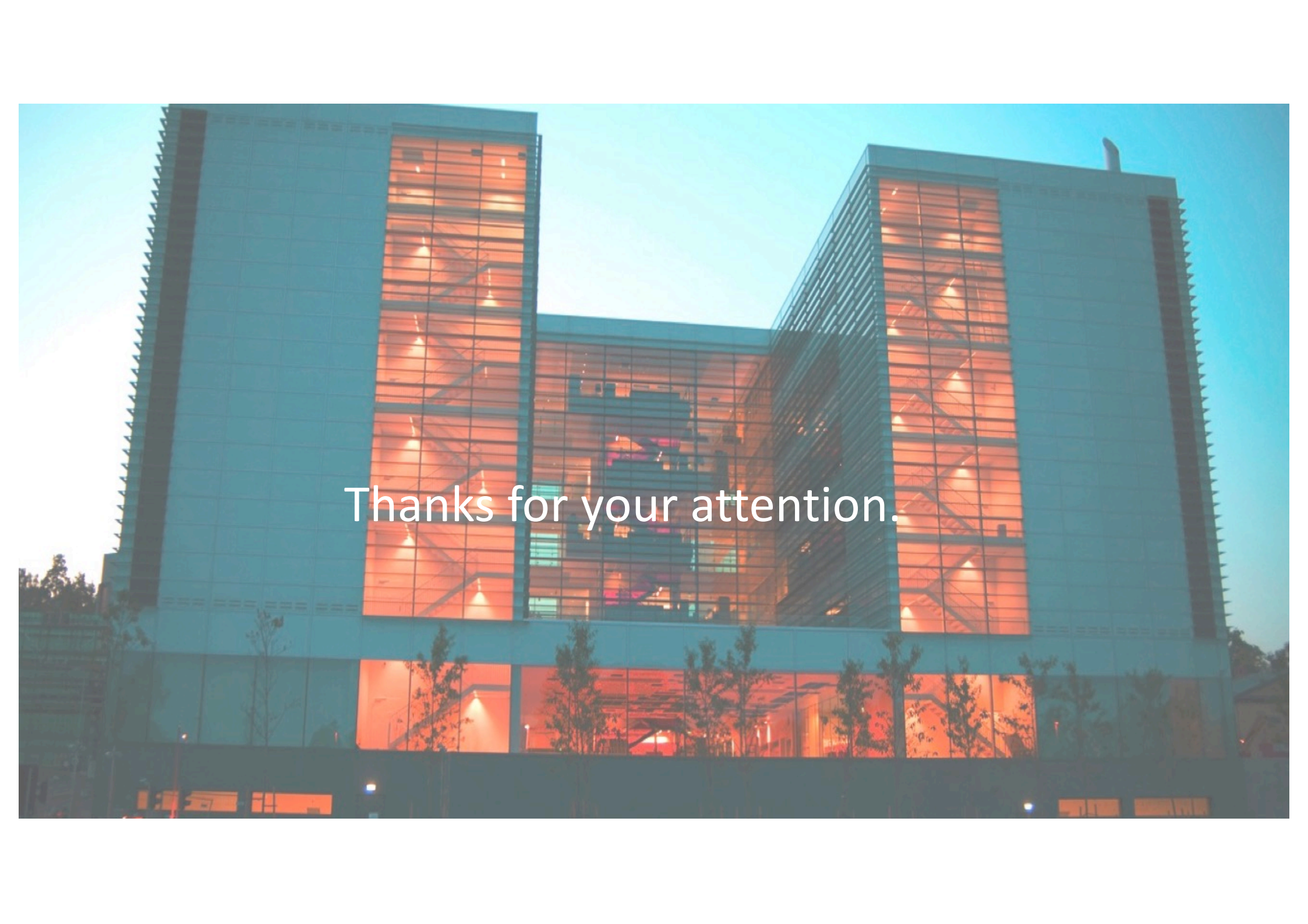


# PCGR DEMO REPORT



## Variant interpretation in cancer: summary

- Comprehensive DNA variant interpretation is critical for implementation of precision cancer medicine
- Types of mutations, number of mutations, mutation locations – all may have therapeutic implications
- Variant consequences are transcript-specific
- A large number of resources have been erected to facilitate clinical interpretation of cancer genomes
- Variant prioritization: tier structure
- Interpretation of the germline background of cancer patients adds an important dimension for clinical translation

A photograph of a modern, multi-story building with a glass facade. The building is illuminated from within, showing a warm orange glow. The central part of the building features a large glass atrium with a staircase. The building is set against a clear blue sky. The text "Thanks for your attention." is overlaid in the center of the image.

Thanks for your attention.